A MAXIMUM LIKELIHOOD APPROACH TO ESTIMATING CORRELATION FUNCTIONS

ERIC JONES BAXTER

Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637.

EDUARDO ROZO

SLAC National Accelerator Laboratory, Menlo Park, CA 94025.
Draft version November 27, 2013

ABSTRACT

We define a Maximum Likelihood (ML for short) estimator for the correlation function, ξ, that uses the same pair counting observables (D, R, DD, DR, RR) as the standard Landy and Szalay (1993, LS for short) estimator. The ML estimator outperforms the LS estimator in that it results in smaller measurement errors at any fixed random point density. Put another way, the ML estimator can reach the same precision as the LS estimator with a significantly smaller random point catalog. Moreover, these gains are achieved without significantly increasing the computational requirements for estimating ξ. We quantify the relative improvement of the ML estimator over the LS estimator, and discuss the regimes under which these improvements are most significant. We present a short guide on how to implement the ML estimator, and emphasize that the code alterations required to switch from a LS to a ML estimator are minimal.

Subject headings: cosmology: large-scale structure of universe

1. INTRODUCTION

While the universe is homogeneous on large scales (e.g. Scrimgeour *et al.* 2012), the galaxies that populate the universe are not distributed uniformly throughout. Rather, galaxies tend to cluster: we are more likely to find a galaxy in a particular patch of the universe if that patch is near another galaxy. The amount of clustering is commonly characterized using the two-point galaxy correlation function, $\xi(r)$, which can be defined as the excess probability relative to the Poisson expectation for a galaxy to be located in a volume element dV at distance r from another galaxy:

$$dP = \bar{n}\,[1 + \xi(r)]\,dV. \tag{1}$$

Here, $\bar{n}$ is the mean number density of galaxies.

The galaxy correlation function is an extremely useful tool in cosmology: it is relatively easy to measure using galaxy surveys (e.g. Davis and Peebles 1983; Hawkins *et al.* 2003; Zehavi *et al.* 2005, and many more), and can be used to estimate cosmological parameters in a variety of ways (for some recent examples see Blake *et al.* 2012; Sánchez *et al.* 2012; Cacciato *et al.* 2012; Tinker *et al.* 2012). With large ongoing and near-future galaxy surveys such as the Baryon Oscillation Spectroscopic Survey (BOSS, Eisenstein *et al.* 2011) and the Dark Energy Survey (DES The Dark Energy Survey Collaboration 2005), among others, it is increasingly important to measure the correlation function quickly and accurately.

The most commonly used methods for determining $\xi(r)$ rely on pair counting. One counts the number of data-data pairs, DD, in the observed galaxy catalog that have some specified radial separation, as well as the number of random-random pairs, RR, in a randomly generated catalog with uniform density and zero correlation. Since ξ quantifies the excess probability for two galaxies to be near each other over the Poisson expectation, one can readily estimate the correlation function via $\hat{\xi} = DD/RR - 1$. More sophisticated estimators

ebaxter@uchicago.edu

have been developed (see Kerscher *et al.* 2000, for a comparison among various estimators), with the most common estimator in employ today being that of Landy and Szalay (1993). The primary advantage of pair counting techniques is that they allow complex survey geometries and masks to be easily dealt with: one simply applies the same mask to both the data and random catalogs when counting pairs.

It has been shown that the correlation function estimator introduced by Landy and Szalay (1993, henceforth LS) is optimal (in the sense that it has the lowest possible variance) in the limit of vanishing correlation function and large data and random catalogs. If these conditions are violated — as they are in the real world, where one is interested in clustered fields and finite catalogs — then it stands to reason that the LS estimator may not be fully optimal.

In this paper we consider the Maximum Likelihood (henceforth ML) estimator for the correlation function (for a similarly minded but much more sophisticated approach towards estimating the power spectrum see Jasche *et al.* 2010). The estimator relies on the same observables as the LS estimator — i.e. D, R, DD, RR, and DR — but, as we will demonstrate, it can achieve greater precision than the LS estimator at the same number of random points. Or equivalently, it obtains identical precision at lower random catalog densities, thus reducing the computation load. We show that our estimator reduces to the LS estimator in the limit that the correlation function vanishes, the survey volume is very large and the catalog densities are large, as one would expect. Our estimator is also very easy to implement (see the summary instructions in §5), so the effort required to switch from a Landy and Szalay (1993) estimator to that advocated in this work is minimal.

Our work bears some similarity to an earlier analysis by Dodelson *et al.* (1997), where they considered the full likelihood function for a galaxy survey, i.e. the probability of finding some set of galaxies at particular positions in a survey. This observable vector can, in principle, contain much more information than the pair counts DD, DR, and RR, so one may expect such an analysis to be superior to ours. How-

ever, as noted in that work, maximizing such a likelihood is not possible in general. Instead, Dodelson *et al.* (1997) found that, in the limit of vanishing clustering, the maximum likelihood estimator reduced to $\hat{\xi}_{ML} = (DD - DR + RR)/DD$, which is very close to the LS estimator. For our purposes, there are two takeaways: first, even though Dodelson *et al.* (1997) considered a much more general problem than that of maximizing the likelihood of the pair-counting observables, their final expression for the correlation function only depends on pair counts in the no clustering limit. Consequently, our analysis should not actually lose any information relative to Dodelson *et al.* (1997) in that limit. The second takeaway is that for clustered galaxy fields, the maximization of the likelihood written down by Dodelson *et al.* (1997) is highly non-trivial. As we discuss below, the maximum likelihood pair counts estimator that we introduce easily accommodates clustering.

The layout of the paper is as follows. In §2 we describe the formalism we use to define the maximum likelihood estimator for $\xi(r)$ from the clustering observables. In §3 we apply our technique to unclustered fields, while §4 presents our results for clustered fields. Our conclusions are given in §5, along with a simple recipe for calculating the maximum likelihood correlation function estimator.

2. CLUSTERING OBSERVABLES AND THE MAXIMUM LIKELIHOOD ESTIMATOR

2.1. *Formalism and Definitions of Observables*

Let n be a homogeneous random field (we will subsequently use n to refer to the number density field of galaxies, having dimensions of $1/$volume). The correlation function of n, denoted ξ, can be defined via

$$\xi(\mathbf{r}) = \frac{\langle n(\mathbf{x})n(\mathbf{x}+\mathbf{r})\rangle - \langle n(\mathbf{x})\rangle\langle n(\mathbf{x}+\mathbf{r})\rangle}{\langle n(\mathbf{x})\rangle\langle n(\mathbf{x}+\mathbf{r})\rangle} \qquad (2)$$

$$= \frac{\langle n(\mathbf{x})n(\mathbf{x}+\mathbf{r})\rangle - \langle n(\mathbf{x})\rangle^2}{\langle n(\mathbf{x})\rangle^2}. \qquad (3)$$

That is, $\xi(\mathbf{r})$ is simply the covariance between any two points separated by a vector $\mathbf{r}$, normalized by the appropriate expectation value. We will further assume the field n is isotropic, so that ξ depends only on the magnitude of the vector $\mathbf{r}$. Our final goal is to estimate the correlation function $\xi(r)$ of n given an empirical point realization of the field. Specifically, given a survey volume V, we assume data points within the survey are a Poisson realization of the random field n.[1]

Traditional clustering estimators such as the LS estimator rely on a set of five observables $\mathbf{x} = \{D, R, DR, DD, RR\}$ from which one may estimate the correlation function ξ. For instance, the LS estimator is given by

$$\hat{\xi}_{LS} = \frac{R(R-1)}{D(D-1)}\frac{DD}{RR} - 2\frac{R-1}{D}\frac{DR}{RR} + 1, \qquad (4)$$

where D is the number of data points within the survey volume of interest, and DD is the number of data pairs within the

[1] Recent work by, for instance, Seljak *et al.* (2009), Hamaus *et al.* (2010) and Baldauf *et al.* (2013) has highlighted the possibility of non-Poisson contributions to the stochasticity of the galaxy and halo fields. These corrections are the result of e.g. halo exclusion. The magnitude of such effects appears to be small (at the few percent level) and their inclusion in the present analysis is beyond the scope of this paper. We note, however, that our framework does not preclude the inclusion of such effects and future work could attempt to study how they modify the maximum likelihood estimator.

radial bin $r \pm \Delta r/2$ at which the correlation function ξ is to be estimated. R and RR are the corresponding quantities for a catalog in which the positions of the data points are chosen randomly; DR is the number of data-random pairs whose separation is in the desired range. The form of the LS estimator presented above differs slightly from the commonly used expression $(DD - 2DR + RR)/RR$. The additional factors of D and R in Eq. 4 allow for data and random catalogs of different number density, while the -1's correct for a small bias in the commonly used estimator owing to the finite size of the catalogs.

Using our model in which data points are obtained from a Poisson random sampling of the density field n we can readily compute the expectation values and covariances (§2.3) of the above observables. We pixelize all space into pixels of volume ΔV such that the density field n is constant within a pixel. Let D_i denote the number of data points in pixel i, which is a Poisson realization of the expectation value $\mu_i = n_i\Delta V$. Since n is homogeneous, the expectation value of μ_i is the same for all pixels, with $\langle\mu\rangle = \bar{n}\Delta V$. The probability distribution for D_i is

$$P(D_i) = \int d\mu_i\, P(D_i|\mu_i)P(\mu_i), \qquad (5)$$

where $P(D_i|\mu_i) = \exp(-\mu_i)\mu_i^{D_i}/D_i!$ is a Poisson distribution with mean μ_i. The first two moments of D_i are

$$\langle D_i\rangle = \langle\mu_i\rangle = \bar{n}\Delta V \qquad (6)$$

$$\langle D_i^2\rangle = \langle\mu_i^2\rangle + \langle\mu_i\rangle = (\bar{n}\Delta V)^2(1+\xi_0) + \bar{n}\Delta V. \qquad (7)$$

where ξ_0 is the correlation function at zero separation.

It is customary to recast this formalism in terms of the density fluctuation

$$\delta_i \equiv \frac{D_i - \langle D_i\rangle}{\langle D_i\rangle}. \qquad (8)$$

By definition, $\langle\delta_i\rangle = 0$, and

$$\langle\delta_i\delta_j\rangle = \xi_{ij} + \delta_{ij}\frac{1}{\bar{n}\Delta V}, \qquad (9)$$

where $\xi_{ij} = \xi(\mathbf{r}_{ij})$ and $\mathbf{r}_{ij}$ is the separation vector between pixels i and j. Eq. 9 is the fundamental building block of our analysis. For future reference, we note that we can rewrite D_i in terms of δ_i via

$$D_i = \bar{n}\Delta V(1+\delta_i). \qquad (10)$$

Note that we have not required that n be a Gaussian random field, only that it be statistically homogeneous.

We are now in a position to define our basic cluster observables. For instance, the total number of data points within the survey volume is the sum

$$D = \sum_i D_i S_i = \bar{n}\sum_i \Delta V(1+\delta_i)S_i, \qquad (11)$$

where S_i is the survey window function, such that $S_i = 1$ if pixel i is in the survey and 0 otherwise. Similarly, we can define the radial weighting function W_{ij} such that $W_{ij} = 1$ if the pixels i and j are separated by a distance $r \in [r - \Delta r/2, r + \Delta r/2]$, and $W_{ij} = 0$ otherwise. The total number of data pairs

in the corresponding radial separation bin is

$$DD = \frac{1}{2} \sum_{ij} D_i D_j W_{ij} S_i S_j \qquad (12)$$

$$= \frac{1}{2} \bar{n}^2 \sum_{ij} (\Delta V)^2 (1 + 2\delta_i + \delta_i \delta_j) W_{ij} S_i S_j. \qquad (13)$$

The expressions for R, RR, and DR are straightforward generalizations of the above formulae.

2.2. *Expectation values*

We now turn to computing the expectation value of our observables. The expectation value for D is

$$\langle D \rangle = \bar{n} \sum \Delta V S_i = \bar{n} V, \qquad (14)$$

where V is the survey volume. Likewise, the expectation value for DD is

$$\langle DD \rangle = \frac{1}{2} \bar{n}^2 \sum_{ij} (\Delta V)^2 \left(1 + \delta_{ij} \frac{1}{\bar{n} \Delta V} + \xi_{ij} \right) W_{ij} S_i S_j. \qquad (15)$$

We can zero out the Poisson term since $\delta_{ij} W_{ij} = 0$. Further, assuming the radial selection W_{ij} is such that $\xi(r)$ is constant within the radial shell of interest, the above expression reduces to

$$\langle DD \rangle = \frac{1}{2} \bar{n}^2 [1 + \xi(r)] \sum_{ij} (\Delta V)^2 W_{ij} S_i S_j. \qquad (16)$$

Defining the volume V_1 such that

$$V V_1 = \sum_{ij} (\Delta V)^2 W_{ij} S_i S_j, \qquad (17)$$

the above expression for $\langle DD \rangle$ can be written as[2]

$$\langle DD(r) \rangle = \frac{1}{2} \bar{n}^2 V V_1 [1 + \xi(r)]. \qquad (18)$$

In the limit that r is much smaller than the survey scale, then $S_i = 1$ will almost certainly imply $S_j = 1$ when $W_{ij} = 1$. Consequently, in the small scale limit,

$$W_{ij} S_i S_j \approx W_{ij} S_i, \qquad (19)$$

and therefore

$$V V_1 \approx \sum_i \Delta V S_i \sum_j \Delta V W_{ij} = V V_{shell}, \qquad (20)$$

where V_{shell} is the volume of the shell over which the correlation function is computed. Note that since $W_{ij} S_i \geq W_{ij} S_i S_j$, this approximation is in fact an upper limit, reflecting the fact that spheres centered near a survey boundary are not entirely contained within the survey window.

The expectation values for the observables R, RR, and DR are readily computed given the above results. We find

$$\langle R \rangle = n_R V \qquad (21)$$

$$\langle DR \rangle = \frac{1}{2} \bar{n} n_R V V_1 \qquad (22)$$

$$\langle RR \rangle = \frac{1}{2} n_R^2 V V_1. \qquad (23)$$

where n_R is the mean density of random points.

2.3. *Covariances*

The covariance matrix between the observables can be computed in a fashion similar to that described above (for a similar approach going directly to ξ, see Sánchez *et al.* 2008). For instance, computing the variance of D, we have

$$D^2 = \sum_{ij} \Delta V^2 \bar{n}^2 (1 + \delta_i)(1 + \delta_j) S_i S_j \qquad (24)$$

$$= \sum_{ij} \Delta V^2 \bar{n}^2 [1 + 2\delta_i + \delta_i \delta_j] S_i S_j. \qquad (25)$$

Note the δ^0 (first) sum reduces to $\langle D \rangle^2$, while the sum that is linear in δ vanishes when we take the expectation value. All that remains is the $\delta_i \delta_j$ term. Using Eq. 9 we arrive at

$$\delta_i \delta_j - \text{term} = \bar{n}^2 \sum_{ij} \Delta V^2 \left(\xi_{ij} + \frac{\delta_{ij}}{\bar{n} \Delta V} \right) S_i S_j \qquad (26)$$

$$= \bar{n} V + \bar{n}^2 \sum_{ij} \Delta V^2 \xi_{ij} S_i S_j. \qquad (27)$$

Putting it all together, we find

$$\text{Var}(D) = \bar{n} V + \bar{n}^2 \sum_{ij} \Delta V^2 \xi_{ij} S_i S_j. \qquad (28)$$

Similar calculations can be performed for the remaining observables and their covariances. Appendix A shows our derivation of the Var(DD) as an example. The total covariance matrix can be expressed as a sum of a Poisson and a clustering contribution,

$$\mathbf{C} = \mathbf{C}_{\text{Poisson}} + \mathbf{C}_{\text{clustering}}. \qquad (29)$$

These are

---

[2] Our expressions are significantly simpler than those in Landy and Szalay (1993). The difference is that Landy and Szalay (1993) hold the number of points within the survey volume fixed, whereas we consider a Poisson sampling of a density field. This both simplifies the analysis, and is the more relevant problem for cosmological investigations. In the limit of a large number of data points, however, these differences become insignificant.

$$\mathbf{C}_{\text{Poisson}} = \begin{pmatrix} \bar{n}V & 0 & \frac{1}{2}\bar{n}n_RVV_1 & \bar{n}^2VV_1 & 0 \\ - & n_RV & \frac{1}{2}\bar{n}n_RVV_1 & 0 & n_R^2VV_1 \\ - & - & \frac{1}{4}\bar{n}n_RVV_1\left[n_RV_2+\bar{n}V_2+1\right] & \frac{1}{2}\bar{n}^2n_RVV_1V_2 & \frac{1}{2}\bar{n}n_R^2VV_1V_2 \\ - & - & - & \bar{n}^2VV_1\left[(\bar{n}V_2)+\frac{1}{2}\right] & 0 \\ - & - & - & - & n_R^2VV_1\left[(n_RV_2)+\frac{1}{2}\right] \end{pmatrix} \tag{30}$$

$$\mathbf{C}_{\text{clustering}} = \begin{pmatrix} \bar{n}^2V^22pt & 0 & \frac{1}{2}\bar{n}^2n_RV_1V^22pt & \bar{n}^3V_1V^22pt+\frac{1}{2}\bar{n}^3V_1V^23pt & 0 \\ - & 0 & 0 & 0 & 0 \\ - & - & -\frac{1}{4}\bar{n}^2n_R^2V_1^2V^22pt+\frac{1}{4}\bar{n}^2n_RV_1^2V2pt_b & \frac{1}{2}\bar{n}^3n_RV_1^2V^22pt+\frac{1}{4}\bar{n}^3n_RV_1^2V^23pt & 0 \\ - & - & - & \bar{n}^4V_1^2V^22pt+\bar{n}^4V_1^2V^23pt+\frac{1}{2}\bar{n}^4V_1^2V^24pt & 0 \\ - & - & - & - & 0 \end{pmatrix}. \tag{31}$$

Our convention for the ordering of the observables is $\mathbf{x} = \{D,R,DR,DD,RR\}$. In the above formulae, we defined V_2, $2pt$, $2pt_b$, $3pt$, and $4pt$ via

$$VV_1V_2 = \sum_{ijk}\Delta V^3W_{ij}W_{jk}S_iS_jS_k, \tag{32}$$

$$V^22pt = \sum_{ij}\Delta V^2\xi_{ij}S_iS_j \tag{33}$$

$$(V_1^2V)2pt_b = \sum_{ijk}\Delta V^3\xi_{ik}W_{ij}W_{kj}S_iS_jS_k \tag{34}$$

$$(V_1V^2)3pt = \left\langle\sum_{ijk}\Delta V^3\delta_i\delta_j\delta_kW_{ij}S_iS_jS_k\right\rangle \tag{35}$$

$$(V_1V)^24pt = \sum_{ijkl}\Delta V^4\xi_{ik}\xi_{jl}W_{ij}W_{kl}S_iS_jS_kS_l$$
$$+\frac{1}{2}\sum_{ijkl}\Delta V^4C^{(4)}_{ijkl}W_{ij}W_{kl}S_iS_jS_kS_l. \tag{36}$$

$C^{(4)}_{ijkl}$ is the fourth order cumulant of the random field, and characterizes the non-gaussian contribution to the 4-point term of the sample variance.

We can derive an upper limit on V_2 in the following way. From Eq. 32, we have

$$VV_1V_2 = \sum_{ij}\Delta V^2W_{ij}S_iS_j\sum_k\Delta VW_{jk}S_k. \tag{37}$$

The sum over k is less than or equal to V_{shell} (regardless of the value of j) so that we have

$$VV_1V_2 \leq \sum_{ij}\Delta V^2W_{ij}S_iS_jV_{shell} = VV_1V_{shell}. \tag{38}$$

Therefore,

$$V_2 \leq V_{shell}, \tag{39}$$

with equality in the limit that the survey boundaries can be ignored (i.e. if the scale of interest is very small compared to the survey volume). We can place a lower limit on V_2 using the fact that the covariance matrix of the R and RR observables must be positive-semidefinite. Enforcing this requirement yields

$$n_R^3V^2V_1\left[n_RV_2+\frac{1}{2}\right]-\left(n_R^2VV_1\right)^2 \geq 0, \tag{40}$$

or

$$n_RV_2 \geq n_RV_1-\frac{1}{2}. \tag{41}$$

Since we can set the value of n_R to be arbitrarily large, we find

$$V_1 < V_2 < V_{shell}. \tag{42}$$

It is more difficult to constrain the terms in $\mathbf{C}_{\text{clustering}}$ as these depend on the details of galaxy clustering. The $2pt$ term can be expressed exactly as

$$2pt = \frac{1}{(2\pi)^3}\int d^3k\, P(\vec{k})|S(\vec{k})|^2, \tag{43}$$

where $P(\vec{k})$ is the galaxy power spectrum and $S(\vec{k})$ is the Fourier transform of the survey window function, i.e.

$$S(k) = \int d^3x\, S(\vec{x})e^{-i\vec{k}\cdot\vec{x}}. \tag{44}$$

If we assume that the galaxy distribution is purely Gaussian, then the three-point function and non-Gaussian contribution to the 4-point functions vanish. If we further consider the limit that the survey volume is very large compared to the scale of interest, we can set $W_{ij}S_iS_j \approx W_{ij}S_i$ and

$$|S(k)|^2 = (2\pi)^3V_{survey}\delta(\vec{k}). \tag{45}$$

In that limit, we have

$$(V_1^2V)2pt_b = \frac{V_{survey}}{(2\pi)^3}\int d^3k\, P(\vec{k})|W(\vec{k})|^2 \tag{46}$$

$$(V_1V)^24pt = \frac{V_{survey}}{(2\pi)^3}\int d^3k\, |P(\vec{k})|^2|W(\vec{k})|^2, \tag{47}$$

where $W(k)$ is the Fourier transform of the radial window function. For a spherical survey with a step radial window function, we have

$$S(k) = 3V_{survey}\frac{j_1(kR_{survey})}{kR_{survey}} \tag{48}$$

$$W(k) = V_{shell}j_0(kR), \tag{49}$$

where R_{survey} is the radius of the spherical survey and R is the radius of the scale of interest. In addition, for the second equation we have assumed that the shell over which the correlation function is computed is thin. Taking the spherical limit allows us to convert the integrals in Eqs. 45, 46, and 47 into one-dimensional integrals over the power spectrum which are straightforward to compute.

We consider now the Poisson contribution to the observable covariance matrix written above in the limit that $n_R \to \infty$. In this limit, we can think of R and RR as having zero variance, so we can solve for both V and VV_1 in terms of R and RR:

$$V = R/n_R \tag{50}$$

$$VV_1 = 2RR/n_R^2. \tag{51}$$

Since R and RR are now fixed, the observable vector reduces to $\mathbf{x} = \{D, DR, DD\}$, and the corresponding covariance matrix is

$$\mathbf{C} = \begin{pmatrix} \bar{n}V & \frac{1}{2}\bar{n}n_R VV_1 & \bar{n}^2 VV_1 \\ — & \frac{1}{4}\bar{n}n_R \dot{V}V_1[(n_R V_2)+1] & \frac{1}{2}\bar{n}^2 n_R VV_1 V_2 \\ — & — & \bar{n}^2 \dot{V}V_1 \left[(\bar{n}V_2)+\frac{1}{2}\right] \end{pmatrix}, \tag{52}$$

where we have ignored the $(\bar{n}V_2)$ term in $\text{Var}(DR)$ since we are assuming $n_R \gg \bar{n}$.

Given our expressions for the means and variances of the observables D, DR and DD and assuming a form for the likelihood function we can evaluate the maximum likelihood estimators for $\bar{n}$ and ξ in this limit. We focus on $\bar{n}$ first. If the only observable is D, and assuming a Poisson likelihood, we arrive at

$$\hat{n} = \frac{D}{V} = n_R \frac{D}{R}. \tag{53}$$

Using a Gaussian likelihood introduces a bias of order $1/\bar{n}V$ arising from the density dependence of the covariance matrix. The above estimator has $\langle \hat{n} \rangle = \bar{n}$ and $\text{Var}(\hat{n}) = \frac{\bar{n}}{V}$. We can perform a similar calculation using only the observable DR. Using a Gaussian likelihood and ignoring the density dependence of the covariance matrix we arrive at

$$\hat{n} = \frac{2DR}{n_R VV_1} = n_R \frac{DR}{RR}. \tag{54}$$

Having treated D and DR as independent observables, we now consider what happens when we adopt a joint treatment. In the limit that the survey volume is very large compared to the scale of interest $V_1, V_2 \to V_{shell}$ and the corresponding covariance matrix takes the form

$$\mathbf{C} = \bar{n}V \begin{pmatrix} 1 & \frac{1}{2}n_R V_{shell} \\ \frac{1}{2}n_R V_{shell} & \frac{1}{4}n_R^2 V_{shell}^2 \end{pmatrix}. \tag{55}$$

This matrix is singular, and its zero eigenvector is $\mathbf{e} = (n_R V_{shell}, -2)$. The corresponding linear combination of observables is

$$e = n_R V_{shell} D - 2DR. \tag{56}$$

Its mean is $\langle e \rangle = 0$, and since $\mathbf{e}$ is a zero eigenvector, $\text{Var}(e) = 0$. In other words, $e = 0$ is a constraint equation that the observables D and DR must satisfy in the large survey limit. Note that neither the mean nor variance of e depend on $\bar{n}$, and therefore the information on $\bar{n}$ is entirely contained in the orthogonal eigenvector.

The orthogonal eigenvector is $\mathbf{e}_\perp = (2, n_R V_{shell})$, corresponding to an observable

$$e_\perp = 2D + n_R \Delta V DR. \tag{57}$$

Its mean is

$$\langle e_\perp \rangle = 2\bar{n}V \left[1 + \frac{1}{2}(n_R V_{shell})^2 \right]. \tag{58}$$

The first term in this sum stems from D, while the second arises from DR. In the limit that $n_R \to \infty$, $n_R V_{shell} \gg 1$, and therefore $e_\perp \approx DR$, so the maximum likelihood estimator becomes that due to DR alone. If $n_R V_{shell} \ll 1$, one has $e_\perp \propto D$, and the joint estimator approaches that due to D alone.

We now turn to estimating ξ, and begin by considering the DR–DD observable subspace in the large survey limit. The corresponding covariance matrix is singular, and is given by

$$\mathbf{C} = \frac{1}{4}\bar{n}^3 VV_1 V_2 \begin{pmatrix} n_R^2/\bar{n}^2 & 2n_R/\bar{n} \\ 2n_R/\bar{n} & 4 \end{pmatrix}. \tag{59}$$

The zero eigenvector is $\mathbf{e} = (-2, n_R/\bar{n})$, corresponding to

$$e = \frac{n_R}{\bar{n}}DD - 2DR. \tag{60}$$

Its expectation value is $\langle e \rangle = \frac{1}{2}n_R\bar{n}VV_1(1+\xi)$ and again $\text{Var}(e) = 0$. Consequently, the equation $e = \langle e \rangle$ is a constraint equation that relates ξ and $\bar{n}$. Explicitly, we have

$$\frac{n_R}{\bar{n}}DD - 2DR = \frac{1}{2}n_R\bar{n}VV_1(\xi - 1). \tag{61}$$

All we need to do now to find the maximum likelihood ξ estimator is to find the corresponding $\bar{n}$ estimator, and insert this in our constraint equation. To do so, we must rely on observables orthogonal to $\mathbf{e}$. Now, consider the following combination of observables which corresponds to an orthogonal eigenvector:

$$e_\perp = 2DD + \frac{n_R}{\bar{n}}DR, \tag{62}$$

which has an expectation value of

$$\langle e_\perp \rangle = \frac{1}{2}\bar{n}^2 VV_1 \left[2(1+\xi) + \frac{n_R^2}{\bar{n}^2} \right]. \tag{63}$$

For $n_R/\bar{n} \gg 1$, the second term dominates, and therefore $e_\perp \approx DR$ in this limit. The second vector orthogonal to $e = (n_R/\bar{n})DD - 2DR$ is D. Thus, D and DR span the space orthogonal to e, and therefore the relevant maximum likelihood estimator for $\bar{n}$ is that discussed earlier. For $n_R V_{shell} \ll 1$, the corresponding estimator is $\hat{n} = D/V$. Replacing into our constraint equation for ξ results in the maximum likelihood estimator

$$\hat{\xi} = \frac{R^2}{D^2}\frac{DD}{RR} - 2\frac{R}{D}\frac{DR}{RR} + 1. \tag{64}$$

This is the Landy–Szalay estimator. Conversely, if $n_R V_{shell} \gg 1$, then the maximum likelihood $\hat{n}$ estimator is that from DR, $\hat{n} = n_R(DR/RR)$, which results in

$$\hat{\xi} = \frac{DD \cdot RR}{DR^2} - 1. \tag{65}$$

This is the Hamilton estimator. Both estimators are recovered in their biased forms, but can easily be corrected to account for this bias. Note too that the bias scales as $1/\bar{n}V$, and therefore vanishes in the limit of infinite data, as it should.

In summary, we see that in the limit that an experiment is Poisson dominated, $n_R \to \infty$, and the survey scale is much larger than the scale of interest, the maximum likelihood estimator for the correlation function is either the Landy–Szalay or the Hamilton estimator. This suggests that neither of these estimators is optimal for realistic surveys with finite size, finite random catalogs and/or clustering. It makes sense, then, to identify the true maximum likelihood estimator to gain a lower variance estimate of the correlation function.

2.5. *The Maximum Likelihood Estimator*

Consider the observable vector $\mathbf{x} = \{D, R, DR, DD, RR\}$. The expectation values of the components of $\mathbf{x}$ and their covariances are given by the equations in the previous sections. We consider $\mathbf{p} = \{\bar{n}, \xi, V, V_1, V_2\}$ to be unknown model parameters. Assuming Gaussian statistics, the likelihood for the parameters $\mathbf{p}$ given an observed data vector $\mathbf{x}$ is

$$\mathcal{L}(\mathbf{p}|\mathbf{x}) \propto \frac{1}{\sqrt{\det \mathbf{C}}} \exp\left(-\frac{1}{2}\left(\mathbf{x} - \langle\mathbf{x}\rangle\right)^T \cdot \mathbf{C}^{-1} \cdot \left(\mathbf{x} - \langle\mathbf{x}\rangle\right)\right),$$
(66)

where $\mathbf{C}$ is the covariance matrix of the observables. The dependence of $\mathcal{L}$ on $\mathbf{p}$ is through $\mathbf{C}$ and $\langle\mathbf{x}\rangle$. The covariance matrix $\mathbf{C}$ can be estimated from data (e.g. using standard jackknife techniques) or from theory (e.g. with simulated data catalogs or by developing a model for galaxy clustering). The ML estimator $\hat{\mathbf{p}}_{ML}$ (which contains the ML estimator for ξ, which we call $\hat{\xi}_{ML}$) is obtained by maximizing the above likelihood with respect to the model parameters $\mathbf{p}$. There are many routes one could take to maximize the likelihood to extract $\hat{\mathbf{p}}_{ML}$ (e.g. brute force, a Newton-Raphson algorithm, etc.); we save discussion of the implementation of such methods for later.

3. ML PERFORMANCE: NO CLUSTERING

We begin by comparing the performance of the ML estimator to the LS estimator on uniform random fields (i.e. $\xi = 0$). As we have seen above, for such fields in the limit that $V, \bar{n}V_1, n_R V_1 \rightarrow \infty$, the LS estimator has minimal variance and is therefore precisely the ML estimator. Here, however, we test the performance of the LS and ML estimators on data sets with finite volume and point densities.

3.1. *ML Performance: Analytic Estimates*

We first wish to determine the relative performance of the LS and ML estimators without making use of any galaxy catalogs (simulated or otherwise). As both LS and ML are unbiased (we checked this explicitly), the relevant quantity for comparing the two estimators is the error on ξ. For the ML estimator, the error on $\hat{\xi}_{ML}$ can be computed using the Fisher matrix. For the Gaussian likelihood we have defined in Eq. 66, the Fisher matrix is given by

$$F_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\right] + \frac{\partial\boldsymbol{\mu}^T}{\partial\mathbf{p}_i}C^{-1}\frac{\partial\boldsymbol{\mu}^T}{\partial\mathbf{p}_j},$$
(67)

where i, j label the components of $\mathbf{p}$, and where commas indicate partial derivatives (e.g. Tegmark 1997). The Fisher matrix is then related to the parameter covariance matrix by

$$\mathbf{F}^{-1} = \mathbf{C}_{param},$$
(68)

where we have used $\mathbf{C}_{param}$ to refer to the covariance matrix of parameters to distinguish it from $\mathbf{C}$, the covariance matrix of observables.

The errors on the LS estimator can be easily computed using propagation of uncertainty. The variance of $\hat{\xi}_{LS}$ is given by

$$\mathrm{var}(\hat{\xi}_{LS}) = \mathbf{J}\mathbf{C}\mathbf{J}^T$$
(69)

where the Jacobian matrix, $\mathbf{J}$, is

$$\mathbf{J} = \left(\frac{\partial\hat{\xi}_{LS}}{\partial D}, \frac{\partial\hat{\xi}_{LS}}{\partial R}, \frac{\partial\hat{\xi}_{LS}}{\partial DR}, \frac{\partial\hat{\xi}_{LS}}{\partial DD}, \frac{\partial\hat{\xi}_{LS}}{\partial RR}\right),$$
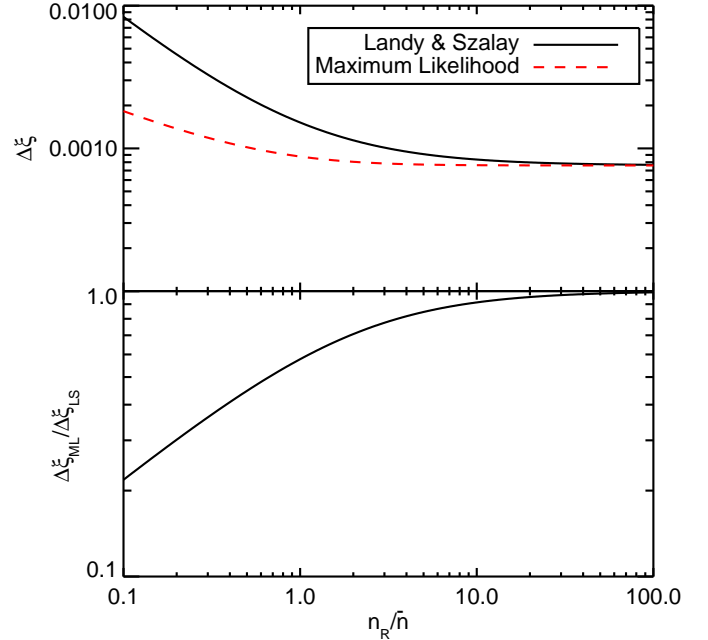(70)

FIG. 1.— *Top panel:* Standard deviation of the ML and LS estimators for ξ as a function of the ratio of the random point density n_R to the data point density $\bar{n}$. *Bottom panel:* The ratio of the standard deviations for the ML and LS estimators, $\Delta\xi_{ML}/\Delta\xi_{LS}$, as a function of $n_R/\bar{n}$. We have assumed a $(1\ h^{-1}Gpc)^3$ survey with $\bar{n} = 5 \times 10^{-5} h^3 \mathrm{Mpc}^{-3}$, $R = 100\ h^3 \mathrm{Mpc}$ and $\Delta R = 10\ h^{-1}\mathrm{Mpc}$. We have set $V_1 = V_2 = V_{shell}$.

and where $\hat{\xi}_{LS}$ is given in Eq. 4. Alternatively, one can derive the above formula by expanding $\hat{\xi}_{LS}$ about its expectation value up to second order in fluctuations, and then evaluating the variance of $\hat{\xi}_{LS}$ in a self-consistent way.

Fig. 1 compares the standard deviations $\Delta\xi_{LS}$ and $\Delta\xi_{ML}$ of the LS and ML estimators as a function of $n_R/\bar{n}$, the ratio of the number density of random points to that of the data points. To make this plot, we hold the input parameter vector, $\mathbf{p}_{input}$, fixed, and vary the random point density as required. We have chosen parameters corresponding to a $(1\ h^{-1}\mathrm{Gpc})^3$ survey with with $\bar{n} = 5 \times 10^{-5}\ h^3 \mathrm{Mpc}^{-3}$, $R = 100\ h^{-1}\mathrm{Mpc}$ and $\Delta R = 10\ h^{-1}\mathrm{Mpc}$. We have also imposed $V_1 = V_2 = V_{shell}$. We see that both the LS and ML estimators converge to the same value of $\Delta\xi$ at large n_R, but that the ML estimator converges much more quickly than the LS estimator.

We now explore how this relative performance depends on the various model parameters. Specifically, looking back at Eq. 30, the covariance matrix depends on four combinations of parameters: $\bar{n}V$, $n_R/\bar{n}$, $n_R V_1$, and V_1/V_2. We find that varying $\bar{n}V$ does not affect the relative performance of LS and ML, so we focus our attention on the remaining three parameter combinations. To further emphasize the difference between the ML and LS estimators, we now focus on the percent "excess error" in $\Delta\xi$ relative to the $n_R = \infty$ value of $\Delta\xi_{ML}$, i.e. we plot

$$\frac{\Delta\xi}{\Delta\xi_{ML}(n_R = \infty)} - 1.$$
(71)

We remind the reader that in the Poisson limit that we are currently considering, both estimators yield the same $\Delta\xi(n_R = \infty)$.

Our results are shown in Fig. 2. The dashed and solid curves show the performance of the LS and ML estimators
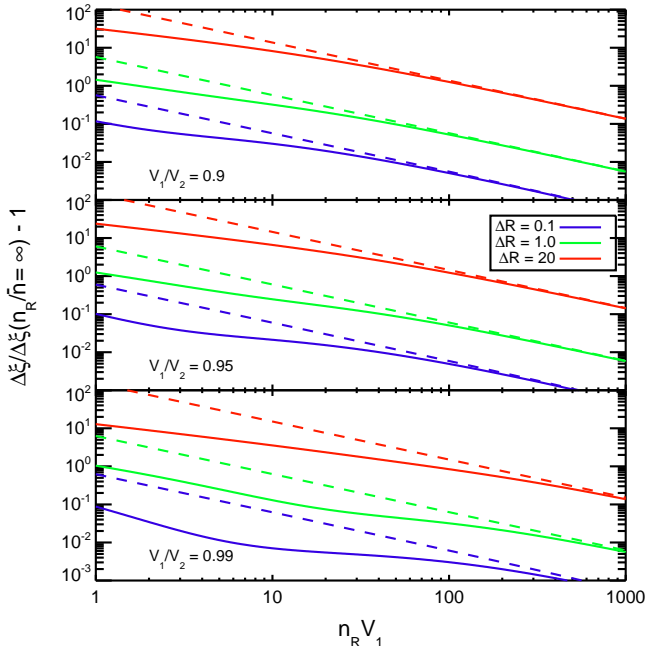
FIG. 2.— The performance of the ML estimator relative to the LS estimator on uniform (Poisson) galaxy fields. Solid lines represent the performance of ML while dashed lines show the performance of LS. We have assume a $(1h^{-1}\mathrm{Gpc})^3$ survey and $R = 100h^{-1}\mathrm{Mpc}$. We have fixed $V_2 = V_{shell}$ (the most conservative assumption for the ML estimator) and show the effect of varying n_R, ΔR and V_1/V_2.
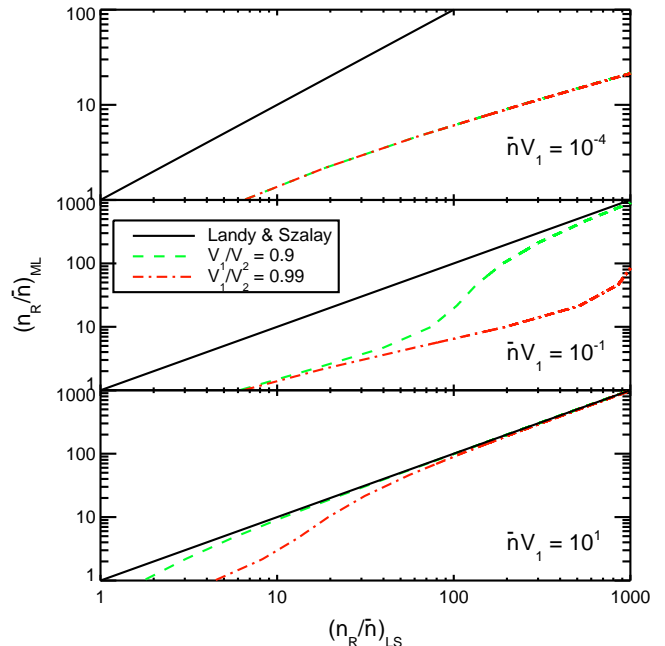

FIG. 3.— The random point density $(n_R/\bar{n})_{ML}$ necessary for the ML estimator to match the precision of the LS estimator with a random point density $(n_R/\bar{n})_{LS}$.

respectively, as a function of n_R, while holding V_1 fixed; i.e. the radial bin used to estimate ξ is fixed. The three sets of curves correspond to three different values for V_1, or equivalently, three different radial bin-widths. Finally, the three panels explore different choices of V_1/V_2. Throughout, we have set $V_2 = V_{shell}$, so that varying V_1/V_2 is equivalent to varying V_1/V_{shell}. We expect this should provide a worst-case scenario for the ML estimator, since the variance of ξ increases with V_2.

Fig. 2 confirms our observation that in the limit that n_R becomes very large, the ML estimator approaches the LS estimator. It can also be seen in the figure that the ML estimator becomes significantly better than LS when $n_R V_{shell} \lesssim 100$, and that this requirement is relatively independent of the other parameters. Likewise, the improvement of the ML estimator relative to the LS estimator is stronger when $V_1/V_2 \approx 1$.

There is an alternative way of viewing the improved performance of the ML estimator that is particularly well suited to a discussion of computational efficiency. Specifically, given an LS estimator with a random point density $(n_R/\bar{n})_{LS}$, one can determine the random point density $(n_R/\bar{n})_{ML}$ required for the ML estimator to achieve the same precision. Fig. 3 shows this ML random point density as a function of the LS random point density. Since typical pair counting algorithms on N points scale as $O(N\sqrt{N})$, this reduction in the number of required random points means that the computation of ξ can be made significantly faster. Because the overall improvement depends on V_2/V_1, we postpone a more quantitative discussion until after we estimate this ratio from numerical simulations below.

As a final note before we turn our attention to numerical simulations, we also found the ML estimator for $\bar{n}$ outper-

forms the standard estimator

$$\hat{\bar{n}} = n_R \frac{D}{R}. \tag{72}$$

Specifically, if $D \lesssim 10^4$ and $(n_R/n_D) \leq 10$, then the ML estimator can outperform the standard estimator by a significant margin. Modern galaxy surveys have $D \gg 10^4$ galaxies, so this result is only significant for estimating the density of rare objects, e.g. galaxy clusters. Note, however, that in that case, it is easy to ensure that $(n_R/\bar{n})$ is very large, so implementing the ML estimator is not necessary. Still, this result was interesting enough we thought it worth mentioning.

3.2. *ML Performance: Numerical simulation*

We have seen above that the ML estimator always performs at least as well as LS, and that in some regimes it performs significantly better. We now address two related questions:

- Are our Fisher matrix results representative of simulated data?

- Do we expect an actual survey to fall in a regime where the ML estimator significantly outperforms the LS estimator? In other words, what values of V_1 and V_2 are characteristic of an actual survey?

We address these questions through numerical simulations.

3.2.1. *Numerical Simulations*

For the $\xi = 0$ case that we are considering at this point, generating a catalog of galaxy positions is trivial. We focus on two possible survey geometries:

1. **Cube**: a cube with side length $1h^{-1}\mathrm{Gpc}$

2. **Survey**: a slab of dimensions $2.24\ h^{-1}\mathrm{Gpc} \times 2.05\ h^{-1}\mathrm{Gpc} \times 0.22\ h^{-1}\mathrm{Gpc}$ from which we have removed a $1\ h^{-1}\mathrm{Mpc} \times 2.05\ h^{-1}\mathrm{Gpc} \times 0.22\ h^{-1}\mathrm{Gpc}$ slice

every 10 h^{-1}Mpc along the longest dimension of the slab. In other words, this survey mask is composed of $\sim$ 200 individual rectangular slabs separated by 1 h^{-1}Mpc gaps.

The cube geometry is far simpler than the survey mask of any realistic survey, while the mask adopted in the survey configuration has far more boundary effects than any real survey is likely to have. Thus, the combination of the two should nicely bracket any real world scenario.

In practice, our catalog is generated in the cubical geometry and then remapped into the slab geometry using the technique of Carlson and White (2010). Although this remapping procedure is unnecessary here as the galaxies are not clustered, it will be important when we subsequently introduce clustering (and it explains the somewhat odd dimensions of our slab).

We perform pair counting on multiple realizations of the simulated catalogs using a kd-tree pair counting algorithm. For illustrative purposes, we consider two different scales:

1. **Large**: 100-101 h^{-1}Mpc

2. **Small**: 2-3 h^{-1}Mpc

These scales are chosen to encompass the wide range over which the correlation function is measured in actual data. At the largest scales, the correlation function is used as a probe of cosmology (e.g. by measuring the BAO feature) while at the smallest scales, the correlation function is used as a probe of galaxy formation and other physics.

3.2.2. Computing V_1 and V_2 on Simulated Catalogs

It is important to accurately estimate V_1 and V_2 because, as shown above, the effectiveness of the ML estimator depends on their values. As V_2 only enters the covariance matrix (and not the mean) of observables, estimating it accurately requires computing the observables over many realizations of survey volume. By contrast, V_1 can be estimated easily by averaging over these realizations: $\hat{V}_1 = 2 \langle RR \rangle / (\bar{n}^2 V)$, where the angled brackets indicate an average over the different realizations. We estimate V_2 by maximizing a likelihood[3]

$$\mathcal{L}\left(V_2|\{\mathbf{x}\}\right) \propto \prod_i^{N_{realizations}} \frac{1}{\sqrt{\det \mathbf{C}(V_2)}}$$
$$\times \exp\left(-\frac{1}{2}\left(\mathbf{x}_i - \langle \mathbf{x} \rangle\right)^T \cdot \mathbf{C}^{-1}\left(V_2\right) \cdot \left(\mathbf{x}_i - \langle \mathbf{x} \rangle\right)\right) \quad (73)$$

where the other parameters have been fixed. When maximizing the likelihood, we enforce the physical requirement that $V_1 < V_2$.

We have have used 700 realizations of the survey volume when computing the best fit values of V_1 and V_2. The results are summarized in Table 1. The first four rows of that table show the values of V_1/V_{shell} and V_2/V_{shell} computed directly from the simulations, while the final two rows show the value of V_1/V_2. For the small scale case, the constraint on V_2/V_{shell} is very noisy owing to the low number of galaxies within the small scale shells. The noise is large enough that our best fit value of V_2/V_{shell} violates the inequality in Eq. 42 (although it is consistent at $\sim 1.4\sigma$). Rather than use this noisy value

---

[3] In practice, just as V_1 can be estimated from RR, V_2 could also be estimated from RRR, i.e. counts of triplets of random points. However, we have chosen not to pursue this possibility.

of V_s/V_{shell} in the results that follow, we have set $V_2 = V_{shell}$ to get the most conservative (lower) limit on V_1/V_2. For the large scale case, the noise is much less and we are able to use the value of V_2 computed from the simulations. As we have seen above, lowering the value of V_1/V_2 worsens the performance of the ML estimator. From Table 1 it is clear that $V_1/V_2 > 0.95$ is a conservative lower limit that should apply even in fairly wild survey geometries.

As discussed above, the important control parameters for the ML estimator are $\bar{n}V_1$ and V_1/V_2. For a survey with a typical number density of $\bar{n} = 5 \times 10^{-5} h^3 \mathrm{Mpc}^{-3}$, our V_1 results correspond to $\bar{n}V_1$ values of roughly 0.004 and 30 for the small and large scales respectively (ignoring the relatively small differences in $\bar{n}V_1$ for the two survey geometries). The value of V_2/V_1 is slightly more difficult to ascertain. At the large scale, we find that $V_1/V_2 \gtrsim 0.95$ for the three survey geometries considered. At small scales, our measurement of V_2 is too noisy to get a good estimate of V_1/V_2. However, we have demonstrated that $V_1 < V_2 < V_{shell}$ so $V_1/V_2 \geq V_1/V_{shell} = 0.986$. Looking back at Fig. 2, we expect the ML estimator to significantly outperform the LS estimator at small scales. At large scales, the value of $\bar{n}V_1$ is large enough that we expect the improvement of LS over ML to be more modest (although still significant for $n_R/\bar{n} \lesssim 100$). Of course, if the width of the large scale shell is reduced so that $n_R V_1$ goes down, we expect ML to begin to significantly outperform LS.

	Cube	Survey
V_1/V_{shell} small scale	0.9965 ± 0.0008	0.9862 ± 0.0007
V_1/V_{shell} large scale	0.8531 ± 0.0006	0.6579 ± 0.0003
V_2/V_{shell} small scale	1.9 ± 0.9	2.3 ± 0.9
V_2/V_{shell} large scale	0.886 ± 0.008	0.689 ± 0.005
V_1/V_2 small scale	$\geq 0.9965 \pm 0.0008$	$\geq 0.9862 \pm 0.0007$
V_1/V_2 large scale	0.963 ± 0.009	0.955 ± 0.008

3.2.3. Computing the Maximum Likelihood Estimator on a Simulated Data Catalog

The likelihood in Eq. 66 depends on the model parameters through both the expectation values of the observables, $\langle \mathbf{x} \rangle$ and the covariance matrix $\mathbf{C}$. However, we have found that we can obtain very accurate results by simply evaluating the covariance matrix for parameters that are reasonably close to $\hat{\mathbf{p}}_{ML}$, and – *keeping the covariance matrix fixed* – vary the parameters in $\langle \mathbf{x} \rangle$ to maximize the likelihood. Our approach simplifies the calculation of $\hat{\mathbf{p}}_{ML}$ significantly so that it reduces to several inversions of a 5×5 matrix[4]. This means that calculating $\hat{\xi}_{ML}$ is not significantly more difficult computationally than calculating $\hat{\xi}_{LS}$.

We note we have explicitly verified that in the Poisson case, the derivatives of $\mathbf{C}$ can be safely neglected. For clustered fields, this is difficult to show in general, but one can make a rough argument. As an illustrative example, consider the 2pt

---

[4] We will refer to the estimator calculated in this way as $\mathbf{p}_{ML}$. Strictly speaking, this estimator differs slightly from the maximum likelihood estimator of the previous section in that we are now fixing the covariance matrix in the likelihood. The differences between the numerical values of the two estimators are negligible, however.

contribution. In the limit of a large survey, we can rewrite Eq. 33 as a sum over radial bins R' with $R' \leq R_{survey}$, so that

$$V^2 2pt = V \sum V_{shell}(R')\xi(R'). \qquad (74)$$

Taking the derivative of $2pt$ with respect to $\xi(R)$ we find

$$\frac{d2pt}{d\xi(R)} = \frac{V_{shell}(R)}{V} \ll 1, \qquad (75)$$

so that, relative to the mean, the information on $\xi(R)$ from the sample variance covariance matrix is always being multiplied by factors of $V_{shell}(R)/V$. Perhaps from a more physical perspective, this can also be argued by noting that the sample variance integrals are dominated by survey-volume scale modes, with small scale modes contributing little because of the filtering by the survey window function.

Computing the maximum likelihood estimator in the fashion described above requires making a choice for the covariance matrix used to analyze the data. We will consider two possibilities for this covariance matrix: (1) the true covariance matrix from which the data is generated, and (2) forming an estimate of the covariance matrix from the data itself (setting $V_2 = V_{shell}$). The first possibility represents the best we could hope to do: we are analyzing the data using the same covariance matrix that was used to generate it.

In the second case, we form an estimate of the covariance matrix from the observed data (i.e. a new covariance matrix for each set of observables) and then compute the maximum likelihood estimator using this covariance matrix estimate. To form the estimate of the covariance matrix, we re-express the Poisson covariance matrix in terms of clustering observables. For instance, since $\langle D \rangle = \bar{n}V$, and $\text{Var}(D) = \bar{n}V$, we simply set $\text{Var}(D) = D$ in the covariance matrix. Similarly, setting R, and RR to their expectation values (see §2.2), we can solve for the various terms that appear in the covariance matrix as a function of the clustering observables D, R, and RR. The exception to this rule is V_2, for which we simply assume $V_2 = V_{shell}$. The full covariance matrix obtained in this way is

$$\mathbf{C} = \begin{pmatrix} D & 0 & \rho RR & 2\rho^2 RR & 0 \\ - & R & \rho RR & 0 & 2RR \\ - & - & \frac{1}{2}\rho RR[N_s(1+\rho)+1] & \rho^2 N_s RR & \rho N_s RR \\ - & - & - & 2\rho^2 RR[\rho N_s + \frac{1}{2}] & 0 \\ - & - & - & - & 2RR[N_s + \frac{1}{2}] \end{pmatrix}, \qquad (76)$$

where we have defined $\rho = D/R$ and $N_s = n_R V_{shell}$. Note that since n_R is known (it is chosen by the observer), the above expression can be computed with no a priori knowledge of the input model parameters $\mathbf{p}_{input}$.

Given one of the above choices for the covariance matrix, we now wish to maximize the likelihood while keeping the covariance matrix fixed. For numerical purposes, it is convenient to reparameterize the parameter space using a new vector $\mathbf{p}' = \{\bar{n}, V, \alpha, \beta\} = \{\bar{n}, V, VV_1, VV_1(1+\xi)\}$ such that the expectation value of the observed data vector, $\langle \mathbf{x} \rangle$, is linear in V, α and β. With this reparameterization, maximizing $\mathcal{L}$ given $\bar{n}$ reduces to a simple matrix inversion problem, so the overall minimum can be easily found using standard 1-dimensional minimization routines.

3.2.4. *Comparing the Numerical and Analytic Calculations of the ML Estimator*

Ideally, to test the two estimators we would generate many simulated data catalogs, perform pair counting on each one,

and compute the corresponding ML and LS estimators. However, pair counting on many catalogs for very high n_R is prohibitively expensive from a computational point of view. Instead, we make a small number of realizations of the survey at reasonable n_R and use these realizations to compute the unknown terms in the covariance matrix – V_1 and V_2 – as described above. We then generate 10^5 Monte Carlo realizations of our clustering observables $\mathbf{x} = \{D, R, DR, DD, RR\}$ by drawing from a multivariate Gaussian with the calibrated covariance matrix. Using our Monte Carlo realizations, we compute the mean and standard deviation of each of our estimators in order to test whether the estimators are unbiased, and the relative precision of the two estimators.

Fig. 4 compares the result of our numerical experiment to the analytic results presented in the last section. The red and black shaded regions represent the results of our numerical experiment for the ML and LS estimators respectively; the width of these regions represents the error on $\Delta\xi$ owing to the finite number of realizations. The solid red and black lines represent the theoretical behavior predicted from the Fisher matrix as described above. We see that the results of our numerical experiment are in good agreement with the results of the Fisher calculation. For this plot we assumed the cubical geometry discussed above; the upper panel corresponds to the small scale, while the lower panel corresponds to the large scale.

The blue line in Fig. 4 shows the ML curve when the covariance matrix is estimated directly from the data using the technique described above. As can be seen, this simple method for estimating the covariance matrix produces results that are as good as the case when the true covariance matrix is exactly known.[5] Thus, we can firmly conclude that in the regimes described above – namely low $n_R V_1$ and $V_1/V_2 \sim 1$ – the ML estimator represents a significantly more powerful tool for estimating the correlation function than the LS estimator.

4. ML PERFORMANCE: CLUSTERED FIELDS

Clustering of galaxies introduces new terms into the covariance matrix of the observables $\mathbf{x} = \{D, R, DR, DD, RR\}$. These new terms involve various integrals of the correlation function and its higher moments over the survey volume (see Eq. 31). Consequently, we no longer expect the LS estimator to be the large volume, large n_R limit of the ML estimator.

Our program here will be very similar to that discussed above for the case with no clustering. The major difference is that in the present case, the covariance matrix of the observables is more difficult to calculate as it depends on the detailed clustering properties of the galaxies. One consequence of this fact is that we cannot easily include the dependence of the covariance matrix on ξ in the Fisher matrix calculation as we did previously. Instead, we will consider the covariance matrix to be fixed. This means that we are throwing out some information, but as we have seen above, including the dependence of the covariance matrix on the model parameters does not significantly affect our constraints on ξ.

4.1. *ML Performance: Analytic Estimates*

As we did previously, we can estimate the errors on our ML estimate of ξ using the Fisher matrix. We set $C_{,i} = 0$ in Eq. 67 as the dependence of C on ξ is not known. To estimate C requires computing the $2pt$, $2pt_b$, $3pt$, and $4pt$ terms.

---

[5] One could imagine too an iterative scheme, where the recovered parameters are used to re-estimate the covariance matrix. As shown in Figure 4, however, this is not necessary.
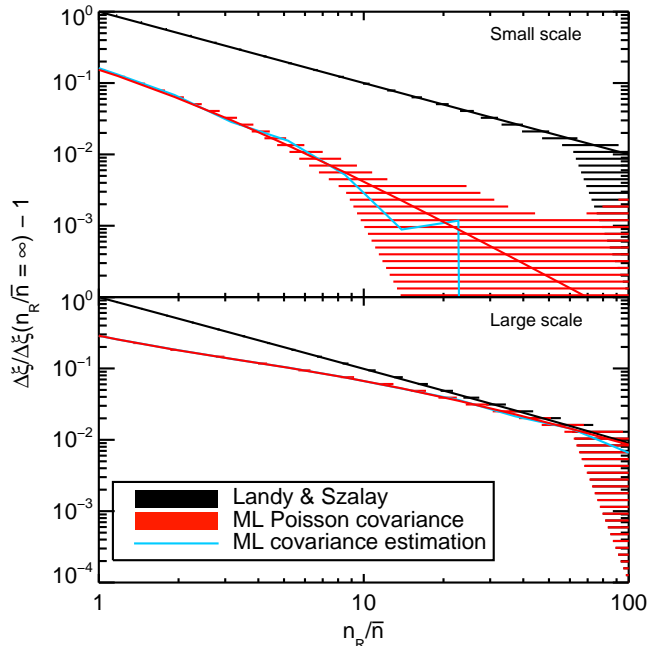
FIG. 4.— Comparison of ML and LS estimators for ξ calculated using simulated galaxy data in the Poisson limit. $\Delta\xi$ is the standard deviation over many random realizations of the observables of the estimator for ξ. The curves have been normalized to the value of the estimator for very large $n_R/\bar{n}$. The shaded regions represent the errors on the measured quantities owing to the finite number of simulations. The solid cuves represent the behavior predicted from the analytic calculations described above. The blue curve corresponds to the ML estimator obtained when expressing the covariance matrix in terms of the clustering observables, as opposed to fixing the covariance matrix to its true value (and setting $V_2 = V_1$). The survey parameters have been chosen to match the cubical survey geometry; upper panel is for the small scale, while the lower panel is for the large scale.
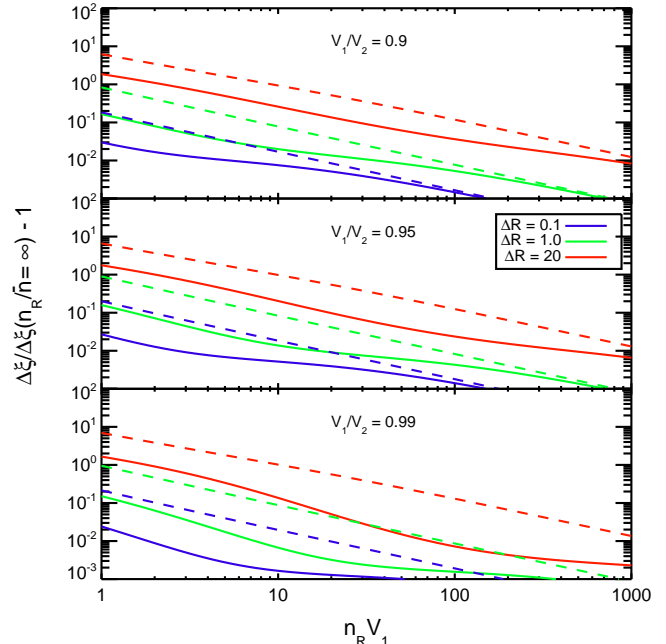
FIG. 5.— The performance of the ML estimator relative to the LS estimator on clustered galaxy fields (analogous to Fig. 2 for unclustered fields). See text for details of our assumptions about cosmology and bias. Solid lines represent the performance of ML while dashed lines show the performance of LS. The errors on ML have been computed using the Fisher matrix, assuming a spherical survey with a volume of $1000\,h^3\mathrm{Mpc}^3$, $\bar{n} = 5 \times 10^{-5}h^3\mathrm{Mpc}^{-3}$, and $R = 100h^{-1}\mathrm{Mpc}$. We have set $V_1 = 0.97V_{shell}$, a resonable value for an actual survey (as illustrated above). We have set $V_2 = V_{shell}$, which gives us the most conservative limit on the performance of the ML estimator.

For our Fisher analysis, we choose to perform the calculation of these terms analytically assuming the Gaussian, large volume, spherical survey limit discussed above. This allows us to express the clustering terms as one-dimensional integrals over the power spectrum (i.e. Eqs. 45, 46, 47, 48, and 49). Therefore, given a power spectrum, we can compute the full covariance of the observables, which in turn allows us to compute the Fisher matrix, and therefore, the error on ξ. We use the power spectrum output from CAMB (Lewis *et al.* 2000) assuming standard ΛCDM cosmological parameters that are consistent with the results of WMAP9 (Bennett *et al.* 2012): $h = 0.7$, $\Omega_c h^2 = 0.1127$, $\Omega_b h^2 = 0.02254$, $n_S = 1.0$. To convert the matter power spectrum into a galaxy power spectrum we have assumed a constant bias of $b = 2$, appropriate for galaxies in a BOSS-like sample.

Fig. 5 shows the results of the Fisher analysis including the effects of galaxy clustering. This figure is analogous to the earlier Fig. 2 which applied to unclustered fields. Comparing the two figures reveals that clustering enhances the performance of ML relative to LS somewhat but that otherwise the qualitative behavior is very similar. The results from our discussion of the ML estimator on unclustered fields therefore carry over to clustered fields mostly unchanged.

4.2. *ML Performance: Numerical Simulation*

As we have done for the unclustered fields, we would now like to connect the results of our Fisher matrix study to results obtained from analyzing simulated realizations of the pair counts observables for clustered galaxy fields. There

are two reasons for going beyond the Fisher marix estimates. First, the simulated realizations allows us to measure $\mathbf{C}_{\text{clustering}}$ in a survey that is more realistic than a sphere; as we will see, this change in geometry significantly impacts the values of the clustering terms. Second, having simulated data allows us to experiment with analyzing this data using different covariance matrices. This is important, as estimating the true observable covariance matrix for clustered galaxy fields is non-trivial.

Measuring the clustering contribution to the covariance matrix requires realizations of a clustered galaxy field. While it is conceivable that the observable covariance matrix could be estimated from a single cosmological realization using a jackknife, we have found that this approach does not yield reliable results at large scales. In the jackknife approach, chunks of the survey volume — which must be significantly larger than the scales of interest — are removed and the pair observables are recalculated; the covariance matrix of the observables can then be related to the covariance across the jackknives. We suspect that the reason this approach does not work in practice is that when dealing with large scales, the size of the removed chunks becomes significant enough that they effectively change the values of V, V_1, and V_2 relative to what they were before each chunk was removed. Consequently, the pair observables computed on the jackknifed survey volume are not drawn from the same underlying covariance matrix as the observables in the full survey volume.

Rather than attempt to address the problems with the jackknives, we instead estimate the observable covariance matrix from multiple realizations of an N-body simulation. We use 41 cosmological realizations of a $(2400\,h^{-1}\mathrm{Mpc})^3$ volume pro-

duced by the LasDamas group (McBride *et al.* 2011). The LasDamas simulations assume a flat ΛCDM cosmology described by $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.04$, $h = 0.7$, $\sigma_8 = 0.8$ and $n_s = 1$. For our "galaxies", we rely on the halo catalog, randomly selecting halos (restricting to $M \geq 10^{13}\,h^{-1}M_\odot$) to achieve the desired number density. Henceforth, we will consider the measurement of the correlation function in a radial bin extending from $R = 50\,h^{-1}\mathrm{Mpc}$ to $R = 60\,h^{-1}\mathrm{Mpc}$. This radial bin was chosen as it is small enough to be significantly impacted by clustering and big enough to contain a large number of galaxies so that the impact of counting noise is minimized.

We estimate the $2pt$, $2pt_b$, $3pt$, and $4pt$ terms in a manner similar to that used to estimate V_2 above. First, we estimate $\bar{n}$, n_R, V and V_1 using the expressions we have derived for the expectation values of the pair counts observables. V_2 is then estimated by maximizing a likelihood as in Eq. 73. Since the D, DR and DD observables are now affected by clustering, however, we consider only the R and RR observables when evaluating the likelihood in Eq. 73. With estimates of $\bar{n}$, n_R, V, V_1 and V_2 in hand, we can form an estimate of $\mathbf{C}_{\mathrm{Poisson}}$. Finally, to determine the clustering terms, we maximize the four dimensional likelihood function defined by

$$\mathcal{L}\left(\{2pt, 2pt_b, 3pt, 4pt\}|\{\mathbf{x}\}\right) \propto$$

$$\prod_{i=1}^{N_{realizations}} \frac{1}{\sqrt{\det \mathbf{C}}} \times \exp\left(-\frac{1}{2}\left(\mathbf{x}_i - \langle\mathbf{x}\rangle\right)^T \cdot \mathbf{C}^{-1} \cdot \left(\mathbf{x}_i - \langle\mathbf{x}\rangle\right)\right)$$

$$(77)$$

where

$$\mathbf{C} = \mathbf{C}_{Poisson} + \mathbf{C}_{clustering}(2pt, 2pt_b, 3pt, 4pt), \qquad (78)$$

and, in our case, $N_{realizations} = 41$. The results of this fitting procedure are shown in Fig. 6. It is clear from the figure that we obtain no significant detection of the $2pt$ and $3pt$ terms. The fact that there is no signficant $3pt$ detection is not surprising as the galaxy field is roughly Gaussian. The non-detection of the $2pt$ term, on the other hand, can be attributed to the fact that we only have 41 realizations of the survey volume and therefore our $2pt$ estimate is noisy. The numerical results for our fits are shown in Table 2.

Although we apparently do not have the constraining power to robustly estimate the $2pt$ term from the N-body simulations, we can estimate it analytically using Eq. 43. For the cubical geometry that we consider here, we can integrate Eq. 43 exactly to obtain an estimate for $2pt$. Performing this calculation requires an estimate of the power spectrum, $P(k)$, which we obtain from CAMB (Lewis *et al.* 2000). The bias is determined by matching the prediction for ξ from the power spectrum to its measured value at the scale of interest; we find that the bias is roughly $b = 1.6$. Our analytic estimate of $2pt$ is shown as a red line in Fig. 6; the numerical value of $2pt = 2.2 \times 10^{-7}$ is in good agreement with the result from the N-body simulations shown in Table 2.

With our estimates of the $2pt_b$, $3pt$, and $4pt$ terms from the fits to the cosmological realizations, and our estimate of $2pt$ by direct integration, we can now compute the full covariance matrix of observables. We use this covariance matrix to generate realizations of the observables as we have done above for the case without galaxy clustering. The results of our analysis of these simulated data sets are presented in Fig. 7. As in Fig. 4, the black curve represents the analytic prediction for the errors obtained using the LS estimator and the shaded region
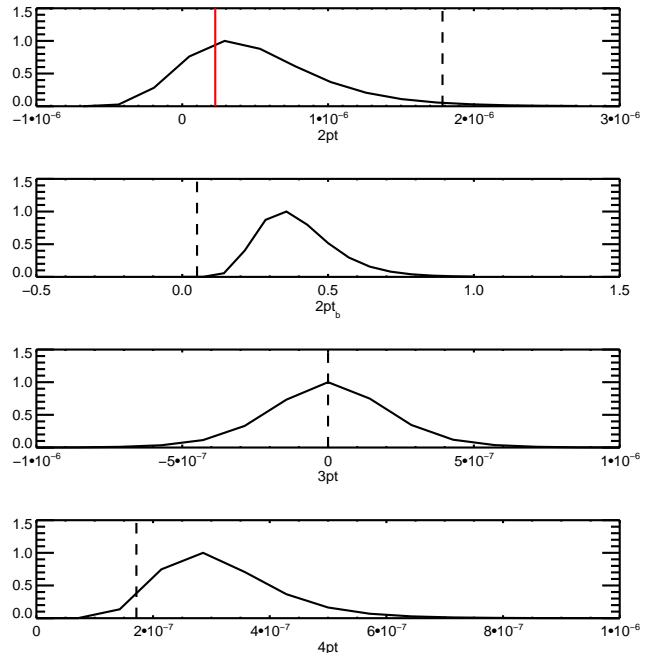

FIG. 6.— The results of our fit for the clustering terms in the covariance matrix to data from N-body simulations. The curve in each panel represents the probability distribution from the fit for the corresponding parameter. The solid vertical line (red in the online version) in the top panel represents our prediction for $2pt$ from integrating Eq. 43 assuming the true cubical geometry of the simulation. Dashed vertical lines in all panels represent predictions for the n-point terms obtained by integrating the expressions we have derived for them assuming the large volume, spherical survey limit (see §2.3).

TABLE 2
FITS TO CLUSTERING TERMS COMPUTED FROM N-BODY SIMULATIONS.

	$R = 50h^{-1}\mathrm{Mpc}$, $\Delta R = 10h^{-1}\mathrm{Mpc}$
$2pt$	$(4 \pm 4) \times 10^{-7}$
$2pt_b$	0.4 ± 0.1
$3pt$	$(0 \pm 6) \times 10^{-8}$
$4pt$	$(3 \pm 1) \times 10^{-7}$

represents the measurement of the errors on simulated data (the width of the region shows the error on this measurement owing to a finite number of realizations). The red curve represents the prediction from the Fisher matrix for the ML estimator, while the red shaded region represents the numerical results computed by analyzing the data using the covariance matrix that was used to generate it.

The analyst wishing to compute the correlation function in a galaxy survey with the ML estimator must first estimate the covariance matrix of D, DR, DD, and RR. In the unclustered case considered previously, the estimation process was straightforward: we simply computed $\bar{n}$, V, and V_1 from the observables D, R, and RR, and we set $V_2 = V_1$, and then substituted into Eq. 30. We showed (blue curve in Fig. 4) that this procedure works well. In the present case, however, computing the observable covariance matrix is more difficult as it depends on the $2pt$, $3pt$, $4pt$, and $2pt_b$ terms, which are not known a priori, and are difficult to estimate from the data.

There are several ways around this difficulty. The simplest is to ignore the clustering contribution to the covariance matrix and simply compute the covariance matrix in the Poisson

limit as above (using only the R and RR observables as these are unaffected by galaxy clustering) for the purposes of defining the ML estimator. Figure 7 compares the performance of this simple approach (blue curve) to the LS estimator and to the ML estimator when run using the correct covariance matrix. It appears that the Poisson covariance matrix approach generally does better than LS but that it does not achieve the maximal performance that can be obtained with the ML estimator using the true covariance matrix. We note that we have also checked that using the Poisson covariance matrix did not bias the resulting ML estimator.

Alternatively, one can use our analytic estimates of the covariance matrix of clustered fields to analyze the data. There are several ways that one could go about this in deail; we take an approach that requires little computational work. As the $2pt$ function can be easily estimated by integrating Eq. 43 for a cubical geometry, we estimate $2pt$ in that way. The remaining clustering terms ($2pt_b$, $3pt$, and $4pt$) are more difficult to estimate in a cubical geomtry, but can easily be estimated for a spherical geometry using Eqs. 46 and 47 (we assume that the distribution is Guassian so that the $3pt$ term vanishes). As seen in Fig. 6, these estimates of the clustering terms are not perfect, but they at least give us some handle on the magnitude of the clustering contribution. Our easy-to-compute estimate of the clustering contribution to the covariance matrix can then be combined with an estimate of the Poisson contribution (as above) to form an estimate of the total covariance matrix. The green curve in Fig. 7 shows the results of analyzing the data using this covariance matrix. We see that this approach generally does better than using only the Poisson covariance matrix, but that it does not do quite as well as using the true covariance matrix.

Finally, one could derive accurate estimates by running several numerical simulations and computing the covariance of the observables across these simulations. Such simulations are typically already perfomed in order to estimate statistical uncertainties, and so this step should not require any additional overhead. This approach can be simulated by using some small number, $N_{realization}$, of the realizations to compute an observable covariance matrix, and then analyzing the data using this covariance matrix. We test this method by using our input covariance matrix to generate $N_{realizations} = 40$ independent data realizations, which in turn are used to estimate the covariance matrix of the generated data. This estimated covariance matrix is then used to estimate ξ. The purple curve in a new independent data set. Fig. 7 shows the errors on ξ calculated in this way. We see that 40 realizations is sufficient to capture the clustering information in the covariance matrix for the purposes of the ML estimator.

One might worry about the circularity of estimating the covariance matrix in the manner described above; after all, the clustering terms in Eq. 31 depend precisely on the correlation function that we are trying to measure. However, as we have seen above, incorrectly estimating the covariance matrix does not lead to a bias in the recovered $\hat{\xi}_{ML}$, but rather increases its variance. Furthermore, even if the clustering terms are set to zero, we still get a lower variance estimate of ξ than with LS. Any reasonable errors in the cosmology used to estimate the covariance matrix will never cause the ML estimator to perform worse than LS. Finally, if an analyst is really worried about obtaining the absolutely minimum variance estimator of ξ, it is always possible to apply the ML estimator in an iterative fashion. One simply assumes a cosmology when cal-
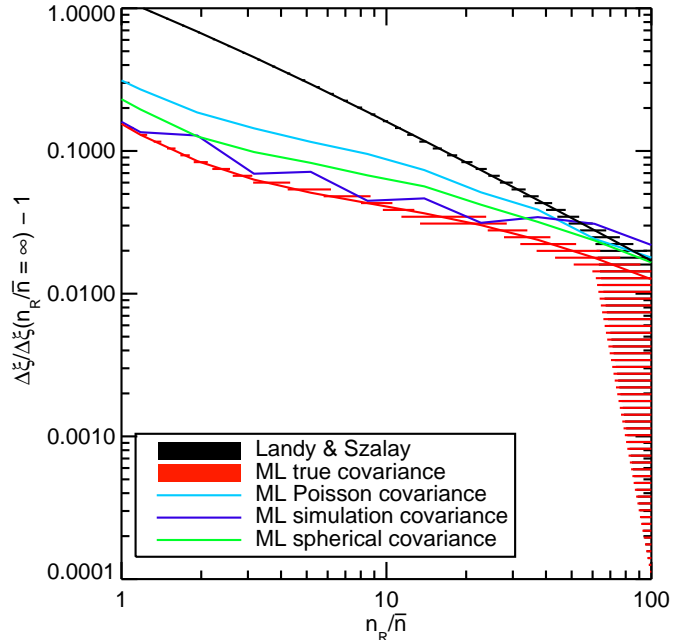

FIG. 7.— The performance of the ML estimator relative to the LS estimator on clustered galaxy fields (analogous to Fig. 4 for unclustered fields). Black and red lines represent analytic predictions for the LS (using Eq. 69) and ML estimator (using the Fisher matrix) respectively. Black and red shaded regions represent the error bands corresponding to these two estimators computed from numerical simulations. The remaining curves represent the results obtained when the simulated data is analyzed with the ML estimator using different choices of the covariance matrix. Blue corresponds to using a Poisson covariance matrix estimated from the data; purple corresponds to using the mean covariance matrix computed from 40 realizations of the survey volume; green corresponds to computing the clustering contribution covariance matrix analytically as described in the text (using the spherical approximation for the $2pt_b$ and $4pt$ terms).

culating the observable covariance matrix, and then adjusts the cosmology based on the recovered $\hat{\xi}_{ML}$; the process can then be repeated until convergence is reached.

5. DISCUSSION

We have explored the utility of the ML estimator for computing galaxy correlation functions. The ML estimator makes use of the same pair-counting observables as the standard LS estimator, yet the former significantly outperforms the latter in certain regimes. Moreover, because all but one of the parameters in the likelihood model are linear (assuming we use a fixed covariance matrix, as described above), the likelihood maximization is numerically trivial. Consequently, we see no reason not to switch from LS estimators to ML estimators: the ML estimator is always better, and has no significant computational requirements in excess of those for the LS estimator. In short, there are only upsides to using the ML estimators, and no real downsides.

For an analyst wishing to compute the correlation function from a galaxy survey, an important question is how large must the random catalog be in order to get errors on ξ that are close to what would be obtained with an infinite random catalog? In Fig. 8 we plot the minimum value of $n_R/\bar{n}$ required to obtain errors on ξ that are within 5% of the value of $\Delta\xi$ at $n_R = \infty$ for both the ML and LS estimators. As we have discussed previously, the ML estimator allows one to compute the correlation function to the same precision as with LS while using a sig-
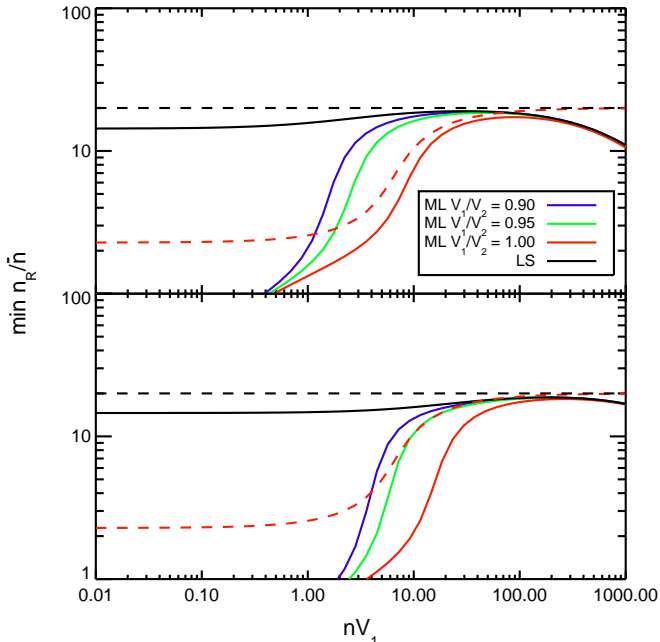
FIG. 8.— The minimum value of n_R required to achieve convergence at the 5% level to the value of $\Delta\xi$ at $n_R = \infty$, as a function of $\bar{n}V_1$. We have assumed a spherical survey with volume $V = (2400 h^{-1}\mathrm{Mpc})^3$. The two panels correspond to different radial scales: upper panel is for $R = 50 h^{-1}\mathrm{Mpc}$ to $60 h^{-2}\mathrm{Mpc}$ while lower panel is for $R = 100 h^{-1}\mathrm{Mpc}$ to $101 h^{-2}\mathrm{Mpc}$. Dashed lines (which do not depend on the radial scale) represent the case where there is no clustering.

nificantly smaller random catalog. We see that LS acheives convergence to the 5% level at $n_R/\bar{n} \sim 15$. Depending on the value of $\bar{n}V_1$, the ML estimator can reduce the required value of n_R by up to a factor of 7.

Perhaps the single biggest obstacle one faces from the point of view of implementing the ML estimator is that one must specify the covariance matrix used to minimize χ^2. In practice, however, we do not believe this is a particularly problematic issue. Firstly, modern cosmological analysis typically rely on extensive numerical simulations to calculate the covariance matrix of survey observables. Just as one can use these numerical simulations to calibrate the variance in $\hat{\xi}_{LS}$, one can use the same simulations to estimate the covariance matrix of the observables D, R, DR, DD, and RR. With this covariance matrix at hand, one can then compute $\hat{\xi}_{ML}$, and use these same simulations to estimate the error $\Delta\hat{\xi}_{ML}$. Alternatively, because the simulation-based covariance matrices are expected to be correct, one could, if desired, treat the problem using Bayesian statistics as usual, without fear of underestimating uncertainties.

Our work can be compared to a recent paper by Vargas-Magaña *et al.* (2013) (hereafter VM). In that work, the authors construct an estimator that significantly outperforms LS on realistic surveys. In brief, their estimator is a linear combination of all possible ratios of the standard pair counts observables $\mathbf{x} = \{DD, DR, RR\}$ up to second order (see their Table 1), i.e.

$$\hat{\xi} = c_0 + \sum_{i=1}^{6} c_i R_i + \sum_{y}^{18} c_i R_i^{(2)}, \qquad (79)$$

where R_i are various ratios of elements of $\mathbf{x}$. The set of coefficients is calibrated using lognormal simulations of the survey with a known correlation function, with the coefficients dependent on the simulations and the survey geometry.

At first sight, an obvious objection to this approach is that because the optimal coefficients depend on the correlation function of the field, the sensitivity of the resulting estimator to choice of correlation function in the log-normal simulations makes this method undesirable. However, VM demonstrated that this problem can be solved with an iterative technique: one uses LS to estimate ξ, and then uses that ξ to generate log-normal realizations, so that the data itself informs the simulations. These realizations are used to define the coefficients, which are then employed in the data to get a new estimate of ξ, and the whole procedure is iterated until convergence is achieved.

When we run our analysis mirroring the random point densities and binning scheme of VM, we find that the improvements of the ML estimator relative to the LS estimator are on the order of several percent, significantly lower than the improvements advertised by VM. Note, however, that Figure 7 demonstrates that the ML estimator saturates the Cramer–Rao bound on the variance of $\hat{\xi}$. As the estimator of VM uses the same pair counts observables as the ML estimator, the fact that the former violates the Cramer–Rao bound may at first appear problematic. The resolution to this problem is that for an arbitrary correlation function, the procedure of VM is biased and therefore the Cramer–Rao bound does not apply. The VM estimator is only unbiased for correlation functions whose shape can be fit by the particular form assumed in their iterative fitting procedure (Eq. 3 of their paper).

As a summary, we would say that if one wishes to quickly and easily estimate an arbitrary correlation function, we can unambiguously advocate the use of the ML estimator over the LS estimator. Under some circumstances, however, where the correlation function is known to be well fit by the form assumed by VM, their iterative scheme leads to a dramatic reduction of errors, at the expense of increased computational requirements and complexity.

5.1. *Recipe for Computing the Maximum Likelihood Correlation Function Estimator*

To aid the reader, we now provide a step-by-step guide of the steps required to implement our ML estimator.

1. Compute the observables $[D, R, DR, DD, RR]$ in the usual fashion.

2. Estimate the covariance matrix of observables:

   - In most cases, we expect this to be done via numerical simulations.
   - One may use a Poisson covariance matrix to analyze the data as in Eq. 76.
   - If desired/necessary, add analytic estimates of the clustering terms to the covariance matrix.

3. Maximize the likelihood defined in Eq. 66 to find $\hat{\xi}_{ML}$, keeping the covariance matrix fixed to the estimate from above. To do so, use the parameter vector $\mathbf{p}' = \{\bar{n}, V, \alpha, \beta\} = \{\bar{n}, V, VV_1, VV_1(1+\xi)\}$, and minimize χ^2. With this redefinition, the only non-linear parameter in our expressions for the expectation values of the

observables (Eqs. 14, 18, 21, 22, 23) is $\bar{n}$. Consequently, minimization can easily be achieved be defining a grid in $\bar{n}$. For each grid point, one finds the maximum likelihood value for the linear parameters through straightforward matrix inversion, and then evaluates the likelihood. The overall minimum can easily be estimated from the data grid.

Before we end, there in one last additional point that is worth noting with regards to correlation function estimators. In particular, our formalism and maximum likelihood framework also suggests what are ideal binning conditions. Specifically, in order to gaurantee that the Gaussian likelihood approximation is good, one should adopt radial bins such that $DD \gg 1$. If one sets $DD = 100$, the corresponding bin width $\Delta \ln R = \Delta R / R$ for a scale R should be

$$\Delta \ln R = 1.59 \left(\frac{1 \text{ Mpc}}{R} \right)^3 \left(\frac{1 \text{ Gpc}^3}{V} \right) \left(\frac{10^{-4} \text{ Mpc}^{-3}}{\bar{n}} \right)^2 . \tag{80}$$

At BAO scales, this suggests that the minimal radial width which one can bin data is therefore $\Delta \ln R \approx 10^{-6}$. This corresponds to exceedingly small angular bins, where the ML estimator is expected to be much superior to the LS estimator. In practice, binning as fine as this is unnecessary, but it does highlight that the ML estimator should enable finer binning than the LS estimator.

REFERENCES

S. D. Landy and A. S. Szalay, ApJ **412**, 64 (1993).
M. I. Scrimgeour *et al.*, MNRAS **425**, 116 (2012), arXiv:1205.6812 [astro-ph.CO].
M. Davis and P. J. E. Peebles, ApJ **267**, 465 (1983).
E. Hawkins *et al.*, MNRAS **346**, 78 (2003), arXiv:astro-ph/0212375.
I. Zehavi *et al.*, ApJ **630**, 1 (2005), arXiv:astro-ph/0408569.
C. Blake *et al.*, ArXiv e-prints (2012), arXiv:1204.3674 [astro-ph.CO].
A. G. Sánchez *et al.*, MNRAS **425**, 415 (2012), arXiv:1203.6616 [astro-ph.CO].
M. Cacciato *et al.*, ArXiv e-prints (2012), arXiv:1207.0503 [astro-ph.CO].
J. L. Tinker *et al.*, ApJ **745**, 16 (2012), arXiv:1104.1635 [astro-ph.CO].
D. J. Eisenstein *et al.*, AJ **142**, 72 (2011), arXiv:1101.1529 [astro-ph.IM].
The Dark Energy Survey Collaboration, ArXiv Astrophysics e-prints (2005), arXiv:astro-ph/0510346.
M. Kerscher, I. Szapudi, and A. S. Szalay, ApJ **535**, L13 (2000), arXiv:astro-ph/9912088.
J. Jasche, F. S. Kitaura, B. D. Wandelt, and T. A. Enßlin, MNRAS **406**, 60 (2010), arXiv:0911.2493 [astro-ph.CO].
S. Dodelson, L. Hui, and A. Jaffe, ArXiv Astrophysics e-prints (1997), arXiv:astro-ph/9712074.
U. Seljak, N. Hamaus, and V. Desjacques, Physical Review Letters **103**, 091303 (2009), arXiv:0904.2963 [astro-ph.CO].
N. Hamaus, U. Seljak, V. Desjacques, R. E. Smith, and T. Baldauf, Phys. Rev. D **82**, 043515 (2010), arXiv:1004.5377 [astro-ph.CO].
T. Baldauf, U. Seljak, R. E. Smith, N. Hamaus, and V. Desjacques, ArXiv e-prints (2013), arXiv:1305.2917 [astro-ph.CO].
A. G. Sánchez, C. M. Baugh, and R. E. Angulo, MNRAS **390**, 1470 (2008), arXiv:0804.0233.
M. Tegmark, Physical Review Letters **79**, 3806 (1997), arXiv:astro-ph/9706198.
J. Carlson and M. White, ApJS **190**, 311 (2010), arXiv:1003.3178 [astro-ph.CO].
A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000), astro-ph/9911177.
C. L. Bennett *et al.*, ArXiv e-prints (2012), arXiv:1212.5225 [astro-ph.CO].
McBride *et al.*, in prep. (2011).
M. Vargas-Magaña, J. E. Bautista, J.-C. Hamilton, N. G. Busca, É. Aubourg, A. Labatie, J.-M. Le Goff, S. Escoffier, M. Manera, C. K. McBride, D. P. Schneider, and C. N. A. Willmer, A&A **554**, A131 (2013), arXiv:1211.6211 [astro-ph.CO]

APPENDIX

DERIVATION OF Var(DD)

We present here a derivation of our expression for Var(DD) in Eqs. 30 and 31. The remaining terms in the covariance matrix can be derived in a similar fashion. We have by definition

$$\text{Var}(DD) = \left\langle (DD(r))^2 \right\rangle - \left\langle DD(r) \right\rangle^2 . \tag{A1}$$

It was shown in the text that

$$\left\langle DD(r) \right\rangle = \frac{1}{2}\bar{n}^2 V V_1 \left[1 + \xi(r) \right]. \tag{A2}$$

Considering the remaining term in Var(DD) and using Eqs. 10 and 12, we find

$$\left\langle (DD(r))^2 \right\rangle = \frac{1}{4}\left\langle \sum_{ijkl} \Delta V^4 \bar{n}^4 (1+\delta_i)(1+\delta_j)(1+\delta_k)(1+\delta_l) W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle \tag{A3}$$

$$= \frac{1}{4}\left\langle \sum_{ijkl} \Delta V^4 \bar{n}^4 (1 + \delta_i\delta_j + \delta_i\delta_k + \delta_i\delta_l + \delta_j\delta_k + \delta_j\delta_l + \delta_k\delta_l + \right. \tag{A4}$$

$$\left. \delta_i\delta_j\delta_k + \delta_i\delta_j\delta_l + \delta_i\delta_k\delta_l + \delta_j\delta_k\delta_l + \delta_i\delta_j\delta_k\delta_l) W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle \tag{A5}$$

$$= \frac{1}{4}\sum_{ijkl} \Delta V^4 \bar{n}^4 \left[1 + \frac{\delta_{ik}}{\bar{n}\Delta V} + \frac{\delta_{il}}{\bar{n}\Delta V} + \frac{\delta_{jk}}{\bar{n}\Delta V} + \frac{\delta_{jl}}{\bar{n}\Delta V} + \xi_{ij} + \xi_{ik} + \xi_{il} + \xi_{jk} + \xi_{jl} + \xi_{kl} \right.$$

$$\left. + \left\langle \delta_i\delta_j\delta_k + \delta_i\delta_j\delta_l + \delta_i\delta_k\delta_l + \delta_j\delta_k\delta_l + \delta_i\delta_j\delta_k\delta_l \right\rangle \right] W_{ij} W_{kl} S_i S_j S_k S_l \tag{A5}$$

$$= \frac{1}{4}\sum_{ijkl} \Delta V^4 \bar{n}^4 W_{ij} W_{kl} S_i S_j S_k S_l + \sum_{ij} \Delta V^3 \bar{n}^3 W_{ij} W_{ik} S_i S_j S_k + \frac{1}{2}\sum_{ij} \Delta V^4 \bar{n}^4 \xi(r) W_{ij} W_{kl} S_i S_j S_k S_l + \bar{n}^4 \sum_{ijkl} \Delta V^4 \xi_{ik} W_{ij} W_{kl} S_i S_j S_k S_l$$

$$+ \bar{n}^4 \left\langle \sum_{ijkl} \Delta V^4 \delta_i\delta_j\delta_k W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle + \frac{1}{4}\bar{n}^4 \left\langle \sum_{ijkl} \Delta V^4 \delta_i\delta_j\delta_k\delta_l W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle \tag{A6}$$

$$= \frac{1}{4}(\bar{n}^2 V V_1)^2 + \bar{n}^3 V V_1 V_2 + \frac{1}{2}\xi(r)(\bar{n}^2 V V_1)^2 + \bar{n}^4 \sum_{ijkl} \Delta V^4 \xi_{ik} W_{ij} W_{kl} S_i S_j S_k S_l$$

$$+ \bar{n}^4 \left\langle \sum_{ijkl} \Delta V^4 \delta_i\delta_j\delta_k W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle + \frac{1}{4}\bar{n}^4 \left\langle \sum_{ijkl} \Delta V^4 \delta_i\delta_j\delta_k\delta_l W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle \tag{A7}$$

Substituting back into the expression for Var(DD) we have

$$\text{Var}(DD) = \bar{n}^3 V V_1 V_2 + \bar{n}^4 \sum_{ijkl} \Delta V^4 \xi_{ik} W_{ij} W_{kl} S_i S_j S_k S_l + \bar{n}^4 \left\langle \sum_{ijkl} \Delta V^4 \delta_i\delta_j\delta_k W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle +$$

$$\frac{1}{4}\bar{n}^4 \left\langle \sum_{ijkl} \Delta V^4 \delta_i\delta_j\delta_k\delta_l W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle - \frac{1}{4}(\bar{n}^2 V V_1 \xi(r))^2 . \tag{A8}$$

The second term in the above expression can be re-written as

$$\sum_{ijkl} \Delta V^4 \xi_{ik} W_{ij} W_{kl} S_i S_j S_k S_l = \sum_{ik} \Delta V^2 \xi_{ik} S_i S_k \sum_j \Delta V W_{ij} S_j \sum_l \Delta V W_{kl} S_l \tag{A9}$$

$$= V_1^2 \sum_{ij} \Delta V^2 \xi_{ij} S_i S_j \tag{A10}$$

$$= V_1^2 V^2 2pt, \tag{A11}$$

where we have used the definition of $2pt$ in Eq. 33. The third term on the right hand side of Eq. A8 can be written as

$$\left\langle \sum_{ijkl} \Delta V^4 \delta_i \delta_j \delta_k W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle = \left\langle \sum_{ijk} \Delta V^3 \delta_i \delta_j \delta_k W_{ij} S_i S_j S_k \sum_l \Delta V W_{kl} S_l \right\rangle \tag{A12}$$

$$= V_1 \left\langle \sum_{ijk} \Delta V^3 \delta_i \delta_j \delta_k W_{ij} S_i S_j S_k \right\rangle \tag{A13}$$

$$= V_1 (V_1 V^2) 3pt, \tag{A14}$$

where we have used the definition of $3pt$ in Eq. 35. Finally, we do a cumulant expansion of the fourth order term to separate out the Gaussian contribution. We have

$$\frac{1}{4} \bar{n}^4 \left\langle \sum_{ijkl} \Delta V^4 \delta_i \delta_j \delta_k \delta_l W_{ij} W_{kl} S_i S_j S_k S_l \right\rangle = \frac{1}{4} \bar{n}^4 \sum_{ijkl} \Delta V^4 \left[\langle \delta_i \delta_j \rangle \langle \delta_k \delta_l \rangle + \langle \delta_i \delta_k \rangle \langle \delta_j \delta_l \rangle + \langle \delta_i \delta_l \rangle \langle \delta_j \delta_k \rangle + C^{(4)}_{ijkl} \right] W_{ij} W_{kl} S_i S_j S_k S_l \tag{A15}$$

$$= \frac{1}{4} \bar{n}^4 \sum_{ijkl} \Delta V^4 \left[\xi_{ij} \xi_{kl} + 2\xi_{ik} \xi_{jl} + 2\frac{\delta_{ik}\delta_{jl}}{\bar{n}^2 \Delta V^2} + C^{(4)}_{ijkl} \right] W_{ij} W_{kl} S_i S_j S_k S_l \tag{A16}$$

$$= \frac{1}{4} \bar{n}^4 (V V_1 \xi(r))^2 + \frac{1}{2} \bar{n}^4 \sum_{ijkl} \Delta V^4 \xi_{ik} \xi_{jl} W_{ij} W_{kl} S_i S_j S_k S_l +$$

$$\frac{1}{2} \bar{n}^2 V V_1 + \frac{1}{4} \bar{n}^4 \sum_{ijkl} \Delta V^4 C^{(4)}_{ijkl} W_{ij} W_{kl} S_i S_j S_k S_l \tag{A17}$$

$$= \frac{1}{4} \bar{n}^4 (V V_1 \xi(r))^2 + \frac{1}{2} \bar{n}^4 (V_1 V)^2 4pt + \frac{1}{2} \bar{n}^2 V V_1, \tag{A18}$$

where we have used the definition of $4pt$ in Eq. 36.

Substituting the above results into our expression for Var(DD) we have

$$\text{Var}(DD) = \left[\bar{n}^3 V V_1 V_2 + \frac{1}{2} \bar{n}^2 V V_1 \right] + \left[\bar{n}^4 (V_1^2 V^2) 2pt + \bar{n}^4 (V_1^2 V^2) 3pt + \frac{1}{2} \bar{n}^4 (V_1^2 V^2) 4pt \right], \tag{A19}$$

in agreement with Eqs. 30 and 31.